

DOCUMENT RESUME

ED 264 255

TM 850 613

AUTHOR Ellett, Chad D; Capie, William
TITLE Assessing Meritorious Teacher Performance: A Differential Validity Study.
PUB DATE Apr 85
NOTE 32p.; Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143) -- Tests/Evaluation Instruments (160)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Classroom Observation Techniques; Data Collection; Elementary School Teachers; Elementary Secondary Education; Master Teachers; *Merit Rating; Peer Evaluation; Secondary School Teachers; State Programs; *Teacher Behavior; *Teacher Effectiveness; *Test Validity
IDENTIFIERS Dade County Public Schools FL; Florida Associate Master Teacher Program; *Teacher Assessment and Development System

ABSTRACT

The Teacher Assessment and Development System (TADS) - Meritorious Teacher Program (MTP) FORM instrument is used in the Dade County Public Schools, Miami, Florida, to evaluate teachers. Its validity for decisions concerning merit pay for master teachers was examined in this study. Specifically, its ability to discriminate between high performing versus other teachers was investigated. Peer teachers nominated superior teachers, and a comparison group was randomly selected. The TADS-MTP form was used to observe 82 teaching behaviors under 19 performance indicators in four categories: knowledge of subject matter; techniques of instruction; classroom management; and student-teacher relationship. The teachers were rated by a trained team including a principal from another school, a supervisor, and a peer teacher. The results indicated that meritorious teachers were nominated by 53.4 percent of their peers. Comparable numbers of elementary, junior, and senior high school teachers were selected. Perfect scores were achieved in 27 of 275 observations. The mean score of the merit teacher group was 67.9, the random teachers, 59.2. This difference was statistically significant. Of the 82 teaching behaviors, 52 significantly differentiated the two groups of teachers, as did the total score. (The list of teaching behaviors is appended.) (GDC)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED264255

Assessing Meritorious Teacher Performance:
A Differential Validity Study

Chad D. Ellett
Louisiana State University

William Capie
University of Georgia

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

C. D. Ellett

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

Paper presented at the annual meeting of the American
Educational Research Association, Chicago, Illinois

April, 1985

TM 850 613

The recent educational reform movement in the United States has stimulated a renewed interest in teacher quality. A central focus of reform has been the improvement of classroom teaching through: 1) restructuring teacher education programs around teacher effectiveness research (Berliner, 1984; Vaughn, 1984); 2) changing teacher certification practices (Hanes and Rowls, 1984); 3) developing more systematic teacher evaluation/assessment programs for making certification, employment and staff development decisions (Wise, Darling-Hammond & McLaughlin, 1984); and 4) changing reward structures and incentives for teacher performance (Schlechty, Joslin, Leak & Hanes, 1984; French, 1984). Of the many reform efforts to improve the quality of teaching, "merit pay" plans and the role of teacher evaluation procedures are perhaps the most controversial (Klein, 1983).

The most visible of large-scale merit pay programs for teachers is the Associate Master Teacher Program currently being implemented in the State of Florida. This program was signed into Florida law in February of 1984. It provides for substantial annual salary increments (up to \$3,000) for teachers who meet educational requirements (graduate degree) and evaluation criteria (passing scores on a knowledge test in the certification field and a classroom observation system). A state-approved system for assessing the classroom performance of teachers for making merit pay decisions in Florida is the focus of this paper.

When the Florida "merit pay" law was passed, a rather comprehensive set of criteria was established by the state for the approval of classroom observation measures to be used in the program. These included rules for conducting classroom assessments of teaching, the scope and sequence of assessment activities and the more traditional psychometric criteria of validity and reliability as well. Currently, only two observation systems have been approved for use in the

Florida merit pay program. These are the Teacher Assessment and Development System (TADS) - Meritorious Teacher Program (MTP) FORM instrument used in the Dade County Public Schools (DCPS), Miami, Florida, and the Florida Performance Measurement System (FPMS) used in all other school systems in Florida. The FPMS was originally developed by a statewide coalition of educators to meet requirements of the Florida Beginning Teacher Program legislated in 1982. The TADS-MTP FORM instrument is an extended version of the standard TADS instrument originally designed in the DCPS to make annual evaluations of teaching. The standard TADS and the TADS - MTP FORM instruments were developed under the supervision of a joint committee representing both management (DCPS) and labor (the United Teachers of Dade/American Federation of Teachers). Development of the standard TADS instrument and a comprehensive observer training program began in 1982. The TADS is being implemented for the first time throughout DCPS (with all 12,000 teachers) during the 1984-1985 school year.

Whenever a measuring instrument is used to identify individuals for any purpose, there is concern about the validity of the decisions to be made. These concerns are not terribly critical when easily measured performance attributes such as swimming speed are assessed. However, when job performance in an area such as teaching is measured to generate data for purposes such as licensure, promotion, dismissal or reward (merit), the validity of the decision becomes a key concern. In these situations, it is not enough to ask whether an instrument is valid or not valid. The research effort must focus on validity questions pertinent to the decisions for which the measuring instrument is designed.

The Florida Associate Master Teacher Program necessitates decisions relative to providing "merit" pay supplements to qualified teachers. Classroom observation instruments are to provide data to assist in making these decisions. Clearly, any instrument developed for use in this program should demonstrate

validity for differentiating between those teacher candidates who possess "merit" attributes and those who do not.

Designing empirical studies to show that an observation instrument possesses adequate differential validity in a merit context is not an easy task. Ideally, the results of a research study would show that the instrument is capable of measuring performance differences between merit and non-merit groups of teachers in a direction consistent with assumptions of the instrument. A fundamental research problem is establishing a reasonable external criterion for identifying "merit" teachers, while at the same time, being careful not to stigmatize teachers in a non-merit comparison group.

Another important concern in these kinds of studies with classroom observation measures is generalizing the validity of the data collection process to assessment parameters and guidelines established for implementing the program. When specified data collection procedures and requirements are not followed, generalization of the results to the decision-making and program operation contexts is considerably weakened. Data collection for research purposes, for example, should be completed under conditions which mirror established program implementation rules to enhance generalization of the results to the program operation context.

Purpose

The purpose of this study was to investigate the differential (discriminant) validity characteristics of the TADS-MTP FORM instrument designed for use in the Florida Associate Master Teacher Program in the Dade County Public Schools (DCPS), Miami, Florida. The results of the study are part of the total research base supporting the validity of the TADS-MTP FORM instrument designed for use in the Florida merit pay program.

Methodology

Establishing the External "Merit" Criterion

A variety of external criteria could conceivably be used to establish a prior classification of teachers as "merit" or "non-merit". Job-related indices of quality teaching (e.g., teacher of the year) or student outcomes such as achievement levels, number of students demonstrating later success, etc. all could be used. However, each of these presents conceptual and practical difficulties, particularly in a school setting like the DCPS where bargaining agreements between management (DCPS) and labor (United Teacher of Dade) (UTD) must be considered in identifying teachers for research purposes.

In this study, the decision was made to allow teachers who otherwise met qualifications for consideration for the Florida Associate Master Teacher Program to nominate from among their qualified colleagues those considered most meritorious. Teachers' nominations were considered to be a practical and reliable means of establishing a high performing teacher group. While individual teachers may lack accuracy in making this judgment, teachers as a group have a collective perspective of variation in the quality of teaching performed by their colleagues within their own school building.

For obvious reasons, a "non-merit" comparison group of teachers could not be selected using the same procedures. Therefore, a random sample of teachers was selected as a comparison group.

Selection of Schools and Teachers

A random sample of 100 schools stratified by school level (Elementary, Junior High and Secondary) was selected from all available schools in DCPS. Nomination lists of teachers eligible for the Florida Associate Master Teacher Program (e.g., years experience, masters degree, etc.) within these schools were developed. Each teacher on this list was told, via mail, to nominate "at least

one but no more than ten percent of the teachers on this list whom you consider to consistently perform at only the highest levels of excellence in the classroom". These nomination data were tabulated and ranked and the fifty teachers receiving proportionately the largest number of nominations for their school levels were selected as a "merit" teacher group. Two teachers who declined to participate were not replaced.

A comparison group of 50 teachers was then randomly selected from within the stratified sample of 100 randomly selected schools. Random sampling procedures were used since, for obvious reasons, it was not feasible to use the teacher nomination procedure to establish a "non-merit" comparison group. Teachers selected were given the option to participate in the study. Only two declined to do so, and these were replaced using the same selection procedures. One teacher who originally declined, later decided to participate. This changed the sample size of the random group to a total of 51 teachers.

Instrumentation

Data for the study were collected with the Teacher Assessment and Development System (TADS)-Meritorious Teacher (MTP) FORM instrument. The TADS - MTP FORM measures 82 teaching behaviors subsumed under 19 performance indicators. The 19 performance indicators define four assessment categories: 1) Knowledge of Subject Matter; 2) Techniques of Instruction; 3) Classroom Management; and 4) Teacher-Student Relations. A copy of the 82 TADS - MTP FORM teaching behaviors classified by assessment categories can be found in APPENDIX A.

A comprehensive (four-day) training program is used to train TADS - MTP FORM observers to proficiency when using the instrument to make merit pay decisions. When scoring the instrument, observers make dichotomous decisions about each of the 82 teaching behaviors in terms of the "acceptability" with which they were performed by the teacher given the nature of the learner, lesson

objectives and the merit performance assessment context. The 82 decisions yield a total instrument score ranging from 0 to 82. This unitary scoring procedure is consistent with the established scoring rules for making merit decisions in the Florida merit pay program.

The content validity of the TADS-MTP FORM instrument has been established using teacher judgements about the instrument items and their use in making merit pay decisions (Ellett & Capie, 1984). Reliability of the instrument and its dependability for making decisions has also been found to be sufficient to meet requirements of the Florida merit pay program (Capie, Ellett & Cronin, 1984). Alternative standards setting procedures for the TADS - MTP FORM have also been explored (Capie, Ellett & Cronin, 1984).

Data Collection Procedures

Classroom observation data were collected on all teachers in the study by teams of trained observers using the TADS-MTP FORM instrument. The Florida State Board Rules for implementing the Associate Master Teacher Program that existed during the spring of 1984 were used to determine the number and kinds of observers to be trained for the study. These rules specified a three-member team consisting of an out-of-building school principal, a supervisor, and a peer teacher. Supervisor observers in this study were DCPS area or district office supervisors and selected representatives of higher education. A total of 61 TADS-MTP FORM observers were trained in two, two-day observer training programs. All of these observers previously had received from two to four days training in the use of the standard TADS instrument. After training, three member assessment teams were established by randomly assigning observers to merit and comparison group teachers within observer "types" until each of the 100 teachers had an assessment team.

All observation data were collected during the late spring, 1984. Each teacher was observed using standardized TADS-MTP FORM procedures for an entire lesson (50 to 55 minutes) by each of the three observers comprising a particular observation team. Observers made independent visits to the classroom and followed procedures consonant with those specified in existing rules for implementing the Florida Associate Master Teacher Program. Data were collected using a single blind procedure since observers did not know whether a particular teacher observed had been previously classified in the "merit" or the random teacher group. A total pool of 275 useable observations was available for analysis.

Data Analysis

Descriptive statistical summaries, proportions of TADS-MTP FORM teaching behaviors credited to teachers in the two groups and probabilities for differences between the groups for each teaching behavior were computed. A discrimination index was also computed by correlating the total number of teaching behaviors demonstrated with each of the individual teaching behaviors.

Results

The results are partitioned into three sections--the results of the selection process, the teaching behavior analyses, and the total score analyses.

Selection Process

Nomination materials were provided to 2,398 teachers in the 100 schools identified. One thousand five hundred and sixty-nine responses (65 percent) were returned. Individual teachers received nominations from zero to one hundred percent of their colleagues who had responded to the nomination form. Although the original intention was to select the 50 teachers with the highest portion of nominations from their colleagues, the plan was modified slightly to

provide a sample which included approximately equal proportions of elementary teachers, junior high and high school teachers. The elementary schools tended to have more consensus as to which teachers were considered "meritorious" than did the larger, secondary schools.

The "merit" teachers were nominated by an average of 53.4 percent of their colleagues. The randomly selected teachers were nominated by an average of 11.3 percent of their colleagues. The range of the two groups was 30 to 100 percent and 0 to 46 percent, respectively. These data are summarized in Table 1.

Inspection of Tables 2 and 3 reveals that the two groups were similar with respect to distribution across school levels and subject areas. The one possible exception is the number of vocational teachers, where one was considered "meritorious" and six were randomly selected.

Teaching Behavior Analyses*

The portion of teachers acceptably demonstrating the various teaching behaviors ranged from 43 percent (IID.4-Lesson is closed appropriately) to 93 percent (IA-Makes no errors indicative of lack of knowledge of subject matter taught, and IIF.2-Attempts are made to clarify confusion which occurs ***or*** no clarification is needed.) (A listing of the 82 TADS-MTP FORM teaching behaviors is included as APPENDIX A). A larger portion of merit teachers demonstrated all teaching behaviors except (IID.3-Non-verbal techniques are used to redirect learners who are persistently off-task ***or*** there is not persistent off-task behavior; and IVB.3-Comments to or about learners are free of demeaning sarcasm and personal ridicule), where the differences between the groups was only 2 percent. The differences in frequencies were statistically

*Numerical/Alphabetical references to specific teaching behaviors can be cross-referenced for content with categories and indicators in the TADS-MTP FORM instrument.

Table 1

Percentage of Merit Nominations
Received by Field Study Participants

Group	Mean	Minimum	Maximum
Merit	53.4	30	100
Random	11.28	0	46

Table 2

Distribution of Field Study
Participants by School Level

Level	"Merit"	Random	Total
Elementary	25	21	46
Junior High	15	15	30
High School	10	15	25
Total	50	51	101

Table 3
Distribution of Field Study
Participants by Subject Taught

Subject	"Merit"	Random	Total
Basic Subjects	25	21	46
Social Science	3	4	7
Natural Science	3	1	4
Fine Arts	2	3	5
Special Education	3	2	5
Language Arts	8	10	18
Mathematics	4	2	6
Physical Education	1	2	3
Vocational	1	6	7
Other			

significant for 52 or 82 behaviors. The largest differences favoring the "merit" group occurred for teaching behaviors (IIB.4-Instructional aids are used skillfully; IIID.1-Stimuli for learners are varied by changing voice, movement, focus of attention, etc.; IIID.3-Non-verbal techniques are used to redirect learners who are persistently off-task ***or*** there is not persistent off-task behavior; and IVB.5-Enthusiasm is communicated for teaching, learning and the subject being taught). At least twenty-five percent more "merit" teachers than random teachers demonstrated these teaching behaviors. These findings are displayed in Table 4.

The total number of teaching behaviors demonstrated was computed for each observation. A correlation (r_{pbi}) calculated between this total and each of the 82 constituent behaviors was used as a discrimination index for the teaching behaviors. The correlation was statistically significant ($p < .05$) in every case, with values ranging from .27 to .66 (teaching behaviors IIB.3-Instructional aids are used at appropriate times in the lesson and IIH.3-Multiple levels of learning are monitored where appropriate). These coefficients are included in Table 4.

Total Teaching Behaviors

"Perfect scores" of 82 were achieved in 27 of the 275 observations (seventeen by merit teachers and 10 by random teachers). None of the teachers averaged 82 teaching behaviors. Scores for single observations ranged to near zero while only two teachers (random group) had averages below 35. These results are displayed in Table 5.

The mean total number of teaching behaviors was 63.5. The mean for the "merit" group was 67.9 while the random group mean was 59.2. The difference in the groups was statistically significant ($p < .01$). The "merit" group was more homogeneous than the random group as shown by the smaller standard deviation for

Table 4
Descriptive Statistics for
TADS-MTP Teaching Behaviors

Teaching Behavior	Grand Mean	"Merit" Group Mean	"Random" Group Mean	Probability of Difference *	Discrimination Index **
I A	93	95	91	.21	.36
IB1	88	94	83	.00	.39
2	69	76	63	.02	.56
3	75	81	69	.01	.58
4	88	90	86	.32	.40
IIA1	83	89	76	.01	.43
2	84	87	82	.30	.44
3	80	84	77	.17	.47
4	69	72	66	.33	.37
5	61	67	54	.03	.47
IIB1	80	85	74	.03	.41
2	76	84	69	.00	.51
3	73	83	64	.00	.53
4	64	76	51	.00	.51
5	49	54	44	.07	.52
IIC1	83	87	79	.10	.35
2	84	87	81	.17	.34
3	83	86	81	.25	.36
4	45	62	50	.04	.43
5	52	57	47	.10	.46
IID1	46	56	36	.00	.59
2	85	90	81	.04	.41
3	85	84	85	.77	.50
4	43	49	36	.04	.45
IIE1	84	90	77	.00	.49
2	88	91	86	.16	.17
3	89	92	87	.20	.42
4	75	79	70	.08	.49
5	73	82	65	.00	.53
IIF1	81	88	74	.00	.44
2	93	97	90	.02	.43
3	85	92	79	.00	.48
4	86	90	81	.03	.42
5	79	86	71	.00	.51

IIG1	89	91	86	.22	.40
2	67	75	59	.00	.56
3	81	86	77	.06	.55
4	50	60	40	.00	.57
IIH1	83	85	80	.26	.41
2	73	81	65	.00	.56
3	60	70	51	.00	.66
4	60	68	52	.01	.43
5	58	64	53	.07	.61
IIII1	72	76	69	.15	.56
2	79	85	72	.01	.40
3	75	82	69	.01	.56
4	65	72	59	.03	.49
IIIA1	93	95	91	.19	.46
2	87	89	86	.43	.49
3	87	90	84	.19	.32
4	80	84	76	.10	.54
IIIB1	87	91	84	.06	.45
2	88	92	85	.08	.49
3	92	92	91	.90	.27
4	89	88	90	.62	.41
IIIC	77	85	70	.00	.60
IIID1	76	89	64	.00	.57
2	71	80	62	.00	.63
3	69	79	49	.00	.58
4	83	90	76	.00	.48
5	72	79	66	.01	.58
6	70	79	61	.00	.59
IIIE1	89	93	85	.05	.48
2	83	88	78	.02	.56
3	74	82	66	.00	.48
4	77	84	69	.00	.51
5	82	86	78	.08	.51
6	80	83	78	.29	.50
IYA1	93	95	91	.19	.36
2	80	85	76	.05	.55
3	75	81	69	.01	.54
4	77	85	69	.00	.53
IVB1	95	96	94	.63	.41
2	87	89	85	.34	.43
3	89	88	90	.62	.32
4	91	95	86	.02	.52
5	65	78	52	.00	.63
6	51	58	44	.03	.43
IVC1	87	92	82	.02	.50

2	92	96	89	.02	.42
3	81	90	72	.00	.52
4	87	91	84	.06	.32
MEAN	63.5	67.9	59.2		

*Probability of χ^2 value associated with group membership and demonstrating behavior.

**Point biserial correlation between behavior and total score.

Table 5
Distribution of Scores Received by
Field Study Participants

Number of Teaching Behaviors Observed	<u>Single Observation</u>		<u>Mean of All Observations</u>	
	<u>"Merit"</u> Teachers	<u>Random</u> Teachers	<u>"Merit"</u> Teachers	<u>Random</u> Teachers
82	17	10	0	0
81	7	4	2	0
80	6	6	1	2
79	4	1	1	2
78	2	2	2	2
77	2	5	1	2
76	7	4	6	0
75	8	3	1	0
74	3	2	2	1
73	5	2	4	2
72	5	2	2	0
71	5	4	1	1
70	0	2	3	2
69	4	3	1	0
68	6	4	1	0
67	3	4	3	1
66	2	1	2	4
65	1	1	1	0
64	2	3	1	1
63	7	1	0	1
62	4	3	2	3

61	5	2	1	1
60	2	3	1	2
59	1	5	1	3
58	0	3	0	1
57	4	3	0	1
56	1	2	1	2
55	1	3	2	0
54	0	3	0	1
53	1	1	0	1
52	1	4	0	2
51	4	2	0	1
50	0	2	0	0
49	2	2	1	0
48	1	2	1	2
47	0	2	1	0
46	0	1	1	0
45	0	1	0	1
44	1	1	0	1
43	1	3	1	0
42	0	1	0	1
41	3	2	0	2
40	1	5	0	0
39	0	4	0	0
38	0	2	0	1
37	0	0	0	1
36	3	2	0	
35	0	2	0	2
Less than 35	2	10	0	2

Table 6

Mean Number of Teaching Behaviors
Demonstrated by Field Study Participants

Group	Single Observations		Mean of All Observations Form	
	Mean	Standard Deviation	Teacher Mean	Standard Deviation
Merit	67.9	12.9	67.7	10.0
Random	59.2	17.1	59.3	13.8
Total	63.5	15.8	63.4	12.7

single observations and for the means of all observations for a teacher. These descriptive statistics are displayed in Table 6.

Discussion

Traditionally, differential (concurrent) validity studies are used to determine if an instrument can discriminate between two groups. Frequently these groups are termed "masters" and "non-masters" according to an external criterion. Interpretation of these results must include an examination of the assumption of different groups used in the study. Clearly, the two groups differed in the number of nominations received. However, the random group was not a "non-masters" group. Instead, it was a diverse group, which apparently included a number of high quality teachers who were not recognized as such by their colleagues. Similarly, the "merit" group may have included teachers who were singled out based on factors such as interpersonal attraction, committee activity, success with special students or other dimensions of professionalism which did not reflect the instructional focus of the Florida Associate Master Teacher Program and the nomination task. In addition, there were several informal "grapevine" reports of teachers colluding to make frivolous nominations to thwart the fidelity of the study as a result of their generally negative attitude toward the very idea of merit pay. The distribution of performance in the merit group confirmed this possibility. Thus, there are numerous considerations to be made in viewing characteristics of the two teacher groups compared in this study.

Despite these sampling limitations, the TADS-MTP FORM was able to discriminate between the two groups. The difference in the mean number of teaching behaviors demonstrated by the two groups was statistically significant ($p < .01$). Inspection of the score distribution in Table 5 shows that scores

favored the merit group. This tendency is more pronounced when the mean for each teacher is used as the unit of analysis. This finding, along with the findings from TADS-MTP FORM reliability analyses (Capie, Ellett, & Cronin, 1984) suggests that means across two or more observations are a more appropriate merit index than data from a single observation. An extended scale would be expected to differentiate the two groups even further.

Of the 82 individual teaching behaviors, 52 significantly ($p < .05$) differentiated the two groups. This finding, at first, appears to be somewhat disappointing. But, it must be tempered by the fact that performance levels were high in a number of cases (e.g., teaching behaviors III B1-3). Nevertheless, even these teaching behaviors did discriminate adequately between low scores and high scores as shown by the 82 discrimination indices.

The differences between the two groups in the performance levels for the individual teaching behaviors were quite revealing. While a few of the 19 TADS-MTP FORM indicators revealed significant differences for all of their constituent teaching behaviors, a more common pattern was to have one or more non-significant behaviors in each indicator. Items which tended to have little discrimination value were behaviors which are generally acknowledged to be desirable and characterized by high performance levels. In general, the behaviors which did differentiate the two teacher groups were areas suggested in the literature as important in differentiating effective and ineffective teachers.

Although few errors in subject matter were found, they were slightly more common in the random teacher group (IA-Makes no errors indicative of lack of knowledge of subject matter taught; and IB.1-Information is up-to-date and timely.) The instruction provided by the meritorious teacher group was considered more appropriate for and relevant to learners (IA-Makes no errors

indicative of lack of knowledge of subject matter taught.) The use of aids (things the teacher uses) (IIB), favored merit teachers more so than did the use of materials (things that learners use) (IIC), perhaps because the use of aids tends to be an optional element in lesson presentation. The merit group had lessons which began in an interesting way and were brought to closure (IID) more often than did the random group, although these behaviors were relatively infrequent in both groups. Explaining (IIE) and clarifying (IIF) were both more effective for the merit group. While acknowledging learners (IIG.1-Learners who try to contribute are acknowledged.) was common for both groups, the remaining three interaction behaviors (IIG.2-Comments, questions, examples, demonstrations and other contributions are sought from learners throughout the lesson; IIG.3-Responses are sufficient to address learners' questions or comments; IIG.4-Learners' ideas are elaborated in the lesson through extended wait-time or teacher comments and/or questions), favored the merit group. Three informal assessment behaviors favored the merit group as did three behaviors relating to the specificity of feedback (II-I.2-Specific feedback is provided to learners about inadequacies in performance; II-I.3-Specific feedback is provided to learners about adequate performances; II-I.4-Suggestions for improving performance are provided to learners).

The first two classroom management indicators represented high performance levels for both groups, perhaps because the observations were prearranged and the study was conducted late in the school year. However, many behaviors in indicators IIIC and IIID showed that merit teachers were more systematic and effective in their management of learner behavior.

Although both groups provided learners with equal opportunities to participate (IVA.1-Learners are provided equal opportunities to participate in class activities), the merit teachers were more supportive of learners as shown

by their scores reflecting reinforcement and encouragement (IVA.2-Learners who respond poorly or who have difficulty are encouraged; IVA.3-Involvement is sought from learners who appear reluctant to actively participate ***or*** there is no necessity for such encouragement; IVA.4-Learners who do well are personally/individually recognized for specific performances.). Similarly, in indicator IVB there was little difference in scores reflecting fairness and patience (IVB.1-Fairness and impartiality are demonstrated when dealing with learners; IVB.2-Patience or empathy or understanding is demonstrated when learners respond poorly or have difficulty; IBV.3-Comments to or about learners are free of demeaning sarcasm and personal ridicule). The merit teachers scored higher than the random teachers on behaviors representing enthusiasm, demonstrating respect, communicating the importance of topics and demonstrating warmth (IVB.3-Comments to are about learners are free of demeaning sarcasm and personal ridicule; IBV.4-Establishes a climate of courtesy and respect; IVB-5-Enthusiasm is communicated for teaching, learning and the subject being taught; IBV.6-The importance of topics to the content area or to real life is stated to learners; and indicator IVC-Demonstrates warmth and friendliness).

There are a number of issues which emerge from this study. They involve methodological and practical concerns, and special concerns in interpreting and applying the results to parameters and rules for implementing Florida's Associate Master Teacher Program and similar programs being contemplated in other states. One of the most novel aspects of the study is the use of teacher nominations as an indication of merit. While the results do not suggest that this method is totally adequate for identifying meritorious teachers, it may be useful when a practical and valid index is needed for comparing groups. Certainly, these results would create suspicions about any process that relied

on nominations alone to identify individuals for merit performance awards or other recognitions.

The results were sufficiently strong to indicate that the TADS-MTP FORM instrument does differentiate between teacher groups which are randomly constituted and those comprised of teachers who are considered meritorious by their teacher colleagues. Given these positive results, the study is also supportive of the data collection methodology, particularly using observers external to the school site. There was some indication in the data that multiple observations discriminated better between the teacher groups than did single observations. This seems reasonable since reliability of the TADS-MTP FORM is increased with multiple observations (Capie, Ellett & Cronin, 1984). Similarly, reliability is enhanced with the amount of observer training. Observers in this study, though experienced with the standard TADS instrument, received only a 2-day training program due to time constraints imposed by the school system. If a full, four-day training session had been completed, differentiation between the merit and random teacher groups may have been even greater.

The study was not designed to determine whether or not one type of observer is more effective than another. All observers in this study were off-site persons, (no teacher was observed by an administrator from their own building), however. There has been some indication that off-site observers are more reliable and more stringent in the application of similar instruments (Capie, Cronin, & Yap, 1985). The number of observations, along with the identity of the observers and their relationship to the teacher candidate, remain the most troublesome policy questions when considering their potential impact on merit decisions and the perceived fairness of programs like Florida's Associate Master Teacher Program. While the TADS-MTP FORM instrument did differentiate the

teacher groups in this study when it was used by external observers, there is no assurance that the results would have been similar had on-site observers (e.g., building principals) been used.

Conclusions

The results of the study were supportive of the abilities of the TADS-MTP FORM instrument to differentiate teachers who were considered meritorious by their teacher colleagues from a group of randomly selected teachers. The greater success of the merit group was statistically significant for 52 teaching behaviors and for the total instrument score. Larger and more frequent differences between these two groups would predictably occur if the groups were "truly" merit and non-merit. Multiple observers were found to be more discriminating between the two groups than single observers. However, the nature of the observers (job titles, etc.) and their relationships to scores for the teachers observed was not investigated.

Considered collectively, the results provide evidence which supports the differential (concurrent) validity of the TADS-MTP FORM for making decisions about meritorious teaching performance in a context which reflects the assessment conditions prescribed by the Florida Associate Master Teacher Program. To the author's knowledge the results of this study are the first to demonstrate that classroom observation measures can be developed to differentiate meritorious teachers who have been identified by an independent criterion (colleague nominations), from other teachers. Demonstrating the validity of instruments like the TADS-MTP FORM relative to the appropriate decision context (merit performance of teachers) can apparently be accomplished. Conducting such studies seems professionally obligatory for policy makers and researchers contemplating, developing, or implementing similar methodologies.

References

- Berliner, D.C. (1984). Making the right changes in preservice teacher education (1984). Phi Delta Kappan, 66, 2, 94-96.
- Capie, W., Cronin, L. and Yap, K.C. (1985). The influence of observer type on assessment decisions and their reliability. Paper presented at the annual meeting of the American Educational Research Association, Chicago, Illinois.
- Capie, W., Ellett, C.D., and Cronin, L. (1985). Assessing meritorious teacher performance: Reliability, decision-making, and standards setting procedures. Paper presented at the annual meeting of the American Educational Research Association, Chicago, Illinois.
- Ellett, C.D. and Capie, W. (1985). The Teacher Assessment and Development System (MTP FORM): Instrument structure, observation procedures and content validation. Paper presented at the annual meeting of the American Educational Research Association, Chicago, Illinois.
- French, R.L. (1985). Dispelling the myths about Tennessee's career ladder program. Educational Leadership, 42, 4, 9-13.
- Hanes, M.L. and Rowls, M.D. (1984). Teacher recertification: A survey of the states. Phi Delta Kappan, 66, 2, 123-126.
- Klein, K. (Ed.) (1984). Merit Pay and Evaluation. Bloomington, Indiana: Center on Evaluation, Development and Research, Phi Delta Kappa.
- Schlechty, P., Joslin, A.W., Leak, S. and Hanes, R.C. (1985). The Charlotte-Mecklenburg teacher career development program. Educational Leadership, 42, 4, 4-8.
- Vaughn, J. (1984). Knowledge resources for improving the content of preservice teacher education. (1984). Journal of Teacher Education, 25, 4, 3-8.
- Wise, A., Darling-Hammond, L. and McLaughlin, M.W. (1984). Teacher Evaluation: A Study of Effective Practices (R-3139-NIE). Santa Monica, Calif: Rand Corporation.

APPENDIX A**List of TADS-MTP FORM Teaching Behaviors**

List of TADS-MTP Teaching Behaviors

CATEGORY I - KNOWLEDGE OF SUBJECT MATTER

- I A1 Makes no errors indicative of lack of knowledge of subject matter taught.
- I B1 Information is up-to-date and timely.
- I B2 Important dimensions or applications of topics are utilized to enhance instruction
- I B3 Subject matter is presented at more than one cognitive or performance level.
- I B4 Sequence of information presented is logical.

CATEGORY II - TECHNIQUES OF INSTRUCTION

- II A1 Instruction is appropriate for the needs and abilities of the learners.
- II A2 Learners have sufficient opportunity to practice lesson objectives.
- II A3 Learners participate in two or more activities which require more than passive listening.
- II A4 The teacher and the learner interact in more than one group size (i.e. class-sized groups, small groups or individual learners) ***or*** the teacher is responsible for only one learner.
- II A5 The lesson is personalized for learners by using the learners' own experiences or by providing examples that are relevant to them.
- II B1 Instructional aids (e.g., chalkboard, pictures, slides, or films, etc.) are appropriate for learners.
- II B2 Instructional aids are appropriate for objectives.
- II B3 Instructional aids are used at appropriate times in the lesson.
- II B4 Instructional aids are used skillfully.
- II B5 Instructional aids enhance accomplishment of instructional objectives.

- II C1 Instructional materials are appropriate for the needs and abilities of the learners.
- II C2 Instructional materials are appropriate for learner objectives.
- II C3 Instructional materials are used at appropriate times in the lesson.
- II C4 Supplemental and/or differentiated materials are used.
- II C5 Instructional materials enhance the accomplishment of lesson objectives.
- II D1 Lesson is initiated with a motivating introduction.
- II D2 Necessary background is or has been established.
- II D3 Instructional components are sequenced in a logical order.
- II D4 Lesson is closed appropriately.
- II E1 Learner attention is ensured before directions and explanations for lesson content are provided.
- II E2 Explanations of lesson content are clear and easy to follow with appropriate vocabulary for learners.
- II E3 Communication is precise with few false starts, interrupters or inappropriate qualifiers.
- II E4 Major points or potential areas of difficulty are emphasized by verbal and/or non-verbal cues and/or by repetition.
- II E5 Examples and/or demonstrations are used to illustrate lesson content.
- II F1 Areas of confusion are identified and communications restated before learners ask questions ***or*** no confusion is evident.
- II F2 Attempts are made to clarify confusion which occurs ***or*** no clarification is needed.
- II F3 Different words and ideas are used in clarification ***or*** no clarification is needed.
- II F4 Clarifications are made for individual learners rather than the entire class when necessary ***or*** no clarification is needed.
- II F5 Attempts to clarify explanations are effective.

- II G1 Learners who try to contribute are acknowledged.
- II G2 Comments, questions, examples, demonstrations and other contributions are sought from learners throughout the lesson.
- II G3 Responses are sufficient to address learners' questions or comments.
- II G4 Learners' ideas are elaborated in the lesson through extended wait-time or teacher comments and/or questions.
- II H1 Monitors learners performance as learners engage in activity.
- II H2 Solicits responses or demonstrations from learners for assessment purposes.
- II H3 Multiple levels of learning are monitored where appropriate.
- II H4 Learners evaluate their own and/or each other's performance.
- II H5 Bases for learner difficulties or misunderstandings are sought ***or*** probing is not necessary.
- II I1 Expectations about learner performance are communicated at the beginning of activities.
- II I2 Specific feedback is provided to learners about inadequacies in performance.
- II I3 Specific feedback is provided to learners about adequate performances.
- II I4 Suggestions for improving performance are provided to learners.

CATEGORY III - CLASSROOM MANAGEMENT

- III A1 Instructional activities begin promptly.
- III A2 There are no unnecessary delays during instruction (e.g., during transitions due to different completion times of group work or during routine tasks).
- III A3 There are no undesirable digressions.
- III A4 Instructional activities fit the allocated time period.
- III B1 Learner attention is ensured before providing directions or explanations for routine tasks.

- III B2 Procedural directions necessary to implement the class activity are clear and complete (e.g., who, what, where, how).
- III B3 Necessary materials are on hand and ready to use.
- III B4 Routine tasks are dealt with in an efficient manner.
- III C1 Approximately 85% or more of the learners are on-task throughout the lesson.
- III D1 Stimuli for learners are varied by changing voice, movement, focus of attention, etc.
- III D2 Active involvement is sought from learners who are only involved passively in instruction ***or*** no learners are passively involved.
- III D3 Non-verbal techniques are used to redirect learners who are persistently off-task ***or*** there is no persistent off-task behavior.
- III D4 Verbal techniques are used to redirect learners who are persistently off-task ***or*** there is no persistent off-task behavior.
- III D5 Efforts to redirect learners who are persistently off-task are successful ***or*** there is no persistent off-task behavior.
- III D6 Techniques are used to maintain the attention of learners who have been redirected ***or*** there is no persistent off-task behavior.
- III E1 Expectations about behavior are made clear to learners ***or*** learner behavior indicates that expectations have been made clear.
- III E2 Consistent expectations about behavior are maintained throughout the lesson.
- III E3 Behavior of the entire class is monitored throughout the lesson.
- III E4 Learners are provided verbal and/or non-verbal feedback about specific behavior(s).
- III E5 Learners who interact inappropriately or otherwise interfere with the work of others are identified and dealt with quickly ***or*** learners do not interfere with instruction.

- III E6 Learners who interact inappropriately or otherwise interfere with the work of others are dealt with appropriately (i.e. firmly and with suitable consequences) ***or*** learners do not interfere with instruction.

CATEGORY IV - TEACHER-STUDENT RELATIONSHIPS

- IV A1 Learners are provided equal opportunities to participate in class activities.
- IV A2 Learners who respond poorly or who have difficulty are encouraged.
- IV A3 Involvement is sought from learners who appear reluctant to actively participate ***or*** there is no necessity for such encouragement.
- IV A4 Learners who do well are personally/individually recognized for specific performances.
- IV B1 Fairness and impartiality are demonstrated with dealing with learners.
- IV B2 Patience or empathy or understanding is demonstrated where learners respond poorly or have difficulty.
- IV B3 Comments to or about learners are free of demeaning sarcasm and personal ridicule.
- IV B4 Establishes a climate of courtesy and respect.
- IV B5 Enthusiasm is communicated for teaching, learning and the subject being taught.
- IV B6 The importance of topics to the content area or to real life is stated to learners.
- IV C1 Warmth and friendliness are demonstrated by a positive tone of voice and eye contact which accompany verbal interaction(s) with learners.
- IV C2 Warmth and friendliness are demonstrated by knowledge and use of student names.
- IV C3 Warmth and friendliness are demonstrated by smiling, laughing or demonstrating a sense of humor.
- IV C4 Warmth and friendliness are demonstrated by sitting or standing near students.